



October 6, 2003

Deployment of the Initial LCG-1 Service at the Fermilab Tier-1 Center

The goal of this document is to provide an overview of the LCG-1 installation process at Fermilab from June 30, 2003 when the first email from the LCG Deployment Area Coordinator, Ian Bird, was received to September 18, 2003 when Fermilab received “official” notification that it had a valid working LCG-1 implementation. A brief overview outlining the perceived expectations, deliverables, and dates for those deliverables will be provided, followed by a detailed description of the technical effort required and difficulties and delays encountered during installation. All work on LCG-0 and LCG-1 was performed by Joe Kaiser of the U.S. CMS Tier-1 Facility Group.

Work Overview and Schedule:

Following an email from Ian Bird on June 30, 2003 Fermilab worked under the understanding that the goal was “to start this deployment in the next two weeks and to have a basic system running at the 10 sites that deployed LCG-0 by the end of July.” A “basic system” at any given site consists of a machine running the following services:

LCFGng Server – Provides the CERN LCG distribution.
UI – User Interface (Users log into this node to run jobs)
CE – Computing Element (This is a “head node”.)
SE – Storage Element (Provides shared storage areas.)
WN – Worker Node (Farm node worker.)

It was determined at the CMS Facilities Meeting that nodes `hotdog45` through `hotdog50` would be allocated for these purposes. This would allow Fermilab to provide a LCFGng server, UI, CE, SE and two worker nodes. Installation work was scheduled to begin in mid-July due to staff vacations. As Fermilab had implemented an LCFGng server with the aid of Marco Serra, initial time estimates indicated that the LCG-1 installation would be complete at the end of July, as per the LCG deployment manager's, or in the first week of August.

It was understood that the schedule of work consisted of the following:

1. Upgrade LCFGng server to the current tagged CVS version.
Create profiles for UI, CE, SE, WN with local configuration parameters.
2. Install CE, SE, WN's via LCFGng or LCFGng lite version.
3. Configure grid information.
4. Configure local batch system.
5. Reinstall with most recent LCG-1.

At the time, sites had the option to use an LCG lite version that did not require the use of

the LCFGng as a RedHat installation server. Fermilab used the full LCFGng server installation to make certain that the LCG-1 could work with the local management schemes and Fermilab security policies. No firm dates for completion or deliverables were indicated at this time.

LCG-1 Deployment

1.LCFGng Upgrade

An LCG-1 installation is a full RedHat 7.3 OS installation with EDG and CERN grid middleware components. To create a proper LCG-1 deployment requires an LCFGng server to install the OS and grid middleware. There is a “lite” version which does not require that a site run an LCFGng server to install the nodes. Technically, the “lite” version should only install LCG-1 grid middleware components that are OS independent. The other, and CERN preferred, option is to use the LCFGng server to install the OS and middleware. This requires a full LCFGng server installation, functioning DHCP, and the ability to PXE boot nodes. Fermilab chose the full LCFGng server implementation.

The process of upgrading the Fermilab LCFGng server began on July 22, 2003. An LCFGng server had been installed with the guidance of Marco Serra in May for the LCG-0 rollout. Upgrading to LCG-1 required checking out a new CVS tag, downloading the most recent RPMS, and updating the repository. This took several days while the system administrator learned about DHCP and learned how to implement DHCP in the Fermi network environment according to Fermilab networking rules, which typically do not allow experiments to run DHCP servers on public subnets. Additionally, the networking group turns on spanning tree algorithms on all switches by default. This causes time delay between a PXE boot request from a node to the DHCP server which will install the node such that the node fails to boot. Nodes failed to install via PXE because of this; however, this took considerable investigation as all other options needed to be ruled out before approaching the networking group to make changes to switch settings. During this investigation, the LCFGng server was reinstalled at least once. This investigation added considerable delay to the LCFGng upgrade.

2.Installation of nodes:

Each node and its configuration is defined in a profile that is turned into an XML document that contains the RPM list to install and the configuration of each given node (UI, CE, SE, WN's). These profiles are built and maintained by the local site and checked into the main LCG-1 CVS tree. CERN provides a basic template for these profiles that local sites are expected to adapt for local needs.

During the LCFGng upgrade process, investigation into DHCP related issues and other difficulties were probed by attempting to install the simplest node type, a worker node or

WN. Profile building and configuring were also explored at this time. Documentation for doing this is provided largely via the LCG email list. Extant documentation may be available but not from any centrally located repository. The attempt to install the worker node allowed us to properly investigate DHCP and network issues that were site specific to Fermi.

One of the critical management requirements is the ability to boot nodes via PXE through the serial console. Due to the complexity and number of machines that we manage, serial console access is necessary to save time and effort. Serial console installation was not available in the LCG-1 installation process. A new boot kernel was built and tested here at Fermilab that allowed for installation via the serial console. This also took a significant amount of time to investigate, troubleshoot, and implement properly.

Additionally, Fermilab has a security environment that requires the use of Kerberos. Kerberos is not built into the LCG-1 and so Fermilab RPMS for openssh need to be integrated into the LCFGng installation server and into each node profile. This is another unexpected step that took extra time and effort to implement and that requires continuing diligence to keep correct.

Various configuration difficulties occurred with each node type and required several installations to arrive at a correct implementation. Included in this process were questions about properly configuring serial console information, turning off unwanted services (kudzu, sendmail, lpd, etc.), partitioning of disks, adding/deleting unwanted RPMS from a node definition (particularly adding Fermi openssh), and mounting needed file systems. These questions were either answered in response to Fermilab posting directly to the LCG-1 rollout list or from information gathered from other sites having similar problems. The main repository for changing configuration header files and rpm.cfg files on the LCFGng server appears to be the rollout list. Joe Kaiser did not find and was not pointed to any central repository for information about editing and configuring this type of information other than the LCG-1 list.

3. Configure grid information:

Each CE and SE required grid certificates and needed to be configured into the CERN site grid resources.

This took several days to acquire, install, and test the certificates. Test jobs were successfully submitted to LCG-1 resources (CERN, RAL) in the last week of August prior to the deployment of LCG1-1_0_0.

4. Configure batch system: PBS is the preferred batch system for the LCG-1 since most sites have a wealth of experience with this product. The initial LCG installation of PBS required the use of a shared file system and the subsequent release required the use of SSH with host based authentication. Fermilab had a number of choices to make with deploying a batch system due to its security policies.

It was requested by the deployment manager that PBS be the batch system due to experiential longevity. USCMS has used FBSNG, a Fermilab batch system product, for use on production and interactive batch systems. Any batch system implemented at Fermi must fit in with the security policies of the lab, which includes kerberos strong authentication. FBSNG is fully integrated with kerberos and allows forwarding and renewing of certificates for jobs. PBS does not have any kerberos authentication mechanisms in it. The choices were to implement FBSNG via the LCFGng mechanism or implement PBS and learn to configure it with Fermi policies. Though we had greater experience with FBSNG, it was decided to implement PBS because integrating it into the LCFGng server mechanism appeared too large an effort with little assurance of success.

There were several possibilities for running PBS on the Fermi LCG-1 deployment:

1. PBS had to be configured with NFS shared file systems. Fermi does not allow for host based authentication.
2. PBS could be configured to use kerberos.

It was decided to pursue option 1 because option 2 required more time and effort than was reasonable. The security team was consulted and theoretical constructs were imagined for doing this properly in PBS, but the time and ability to implement were in doubt with the manpower on hand.

Option 1 was pursued with due diligence; however, communication with the LCG rollout list was required as it was unclear how to revert the PBS configuration back to use NFS mounted file systems, at least to a novice. After significant amounts of mail and help were given via the email list, PBS was finally configured correctly.

5. Reinstall everything with LCG1-1_0_1

LCG-1 is in flux and has had to fix on releases to install. The LCG1-1_0_1 official release was available on September 2 and all sites were asked to install this version.

LCG1_1_0_0 was installed on Fermi resources. This required going back and reconfiguring configuration files whose format had changed, making PBS use shared file systems rather than the default SSH configuration, reinstalling all nodes, and updating the repository.

6. Reinstall everything with LCG1-1_0_1

Several critical vulnerabilities were discovered in openssh, send mail, and pine, all of which run natively on LCG-1 nodes. These were integrated into the LCG-1-1_0_1 and the deployment manager asked that we upgrade.

The LCFGng server was updated, profiles recreated, nodes reinstalled, PBS reconfigured, header files and rpm con fig files recreated and then everything was reinstalled. This was completed on September 19. A request was made to run jobs at the Fermilab site and this was successful.

A brief history of LCG-1 Deployment Time:

July 8, 2003 – Response to initial LCG-1 questionnaire as distributed by deployment manager.

July 10, 2003 – July 22, 2003: Vacation.

July 22, 2003 – August 1, 2003: Return from vacation. LCFGng/DHCP configuration and installation. LCFGng server installation complete by August 1. DHCP functioning, nodes are installing though configuration is in experimental stages.

August 4, 2003 – August 8: Unexpected power outage at Fermilab with week-long fallout. Only one available CMS system administrator for August 4 and 5. Serial-console kernel built, vetted, and installed on August 8.

August 11 – August 15, 2003: Continued configuration and installation of CE, SE, and WN's. Various and sundry changes to installation parameters made. Discussion with Ian Bird and Michael Ernst with current status of LCG-1 at Fermilab. Vacation August 15.

August 18 – 22, 2003: Nodes are installed. Applied for appropriate grid certificates. Tested ability to send jobs to CERN grid resources. Vacation August 22 – 26.

August 27 – 29, 2003: Fermilab work, only one CMS system administrator available. Continuing node configuration effort on LCG-1. Stuck on PBS.

September 2 – 5, 2003: LCG1-1_0_0 released. Re-update LCFGng server, reinstall all nodes, make configuration changes to header files and rpm cfg files, attempts to configure PBS. Sent email to list for help.

September 8 – 12, 2003: Continuing PBS problems.

September 15 – 19, 2003: Fixed PBS problems, various rpm install problems. Testing jobs on local and regional sites, problems with RAL resource broker makes troubleshooting of Fermi installation difficult.

September 22 – 26, 2003: Updated LCFGng server to LCG1-1_0_1 with security updates, upgraded nodes. Ran local test jobs. CERN ran grid test jobs which were successful. Updated status web pages. Checked FNAL profile templates into CVS.

Communications:

A search on “jlkaiser” in the LCG rollout list shows that there has been a continuum of communication on the LCG-1 rollout list from Fermilab. There is a brief period in August (August 15 – September 02) where there are no postings on the list but these are

consistent with vacation, time spent configuring local parameters, and the rollout of a new version of the LCG-1. Additionally, since this is a collaborative exercise, many questions were answered from others on the list.

Impressions:

One of the jobs of the systems administrator is to test, verify, and give assurances that any addition to a system, whether it be hardware or software, is in compliance with the policies and standards of the site. This effort takes time as one must understand the nature of the new system to be integrated, the nature of how it will affect existing systems, and the nature in which it can be easily managed in conjunction with the pre-existing system. LCG-1 provided an interesting use case as it did not fit natively into Fermilab's security policy, was not easily managed from the desktop, and had primitive versions of installation and management mechanisms all ready in use.

The effort required to do the job properly given the constraints of time and effort allocated was far greater than estimated by the LCG-1. The deployment could have been greatly improved if there were definable dates for well-defined deliverables so that effort could be properly allocated at local sites.

In general, members of the list were supportive of the effort expended to install an LCG-1 site. To U.S. CMS, there was no indication that the pace of the effort or commitment of resources was lacking, and such feedback was also not given to those actually deploying LCG-1.